UC Irvine
UC Irvine Previously Published Works

Title
Artificial Intelligence in Gastrointestinal Endoscopy

Permalink
https://escholarship.org/uc/item/5wc764vh

Journal
Clinical Endoscopy, 53(2)

ISSN
2234-2400

Authors
Abadir, Alexander P
Ali, Mohammed Fahad
Karnes, William
et al.

Publication Date
2020-03-30

DOI
10.5946/ce.2020.038

Peer reviewed

Artificial Intelligence in Gastrointestinal Endoscopy

Alexander P. Abadir[1], Mohammed Fahad Ali[1], William Karnes[2] and Jason B. Samarasena[2]

[1]Department of Medicine, University of California Irvine, Orange, CA, [2]Division of Gastroenterology & Hepatology, Department of Medicine, H. H. Chao Comprehensive Digestive Disease Center, University of California Irvine, Orange, CA, USA

Artificial intelligence (AI) is rapidly integrating into modern technology and clinical practice. Although in its nascency, AI has become a hot topic of investigation for applications in clinical practice. Multiple fields of medicine have embraced the possibility of a future with AI assisting in diagnosis and pathology applications.

In the field of gastroenterology, AI has been studied as a tool to assist in risk stratification, diagnosis, and pathologic identification. Specifically, AI has become of great interest in endoscopy as a technology with substantial potential to revolutionize the practice of a modern gastroenterologist. From cancer screening to automated report generation, AI has touched upon all aspects of modern endoscopy.

Here, we review landmark AI developments in endoscopy. Starting with broad definitions to develop understanding, we will summarize the current state of AI research and its potential applications. With innovation developing rapidly, this article touches upon the remarkable advances in AI-assisted endoscopy since its initial evaluation at the turn of the millennium, and the potential impact these AI models may have on the modern clinical practice. As with any discussion of new technology, its limitations must also be understood to apply clinical AI tools successfully. **Clin Endosc 2020;53:132-141**

Key Words: Artificial intelligence; Colonoscopy; Computer assisted diagnosis; Early detection of cancer; Endoscopy

INTRODUCTION

Artificial Intelligence (AI) is becoming rapidly integrated into modern technology in various specialties and industries, including gaming, physics, weather, and social networking.[1,2] In medicine, AI is already being used to distinguish malignant melanomas from benign nevi and to identify diabetic retinopathy and diabetic macular edema from retinal fundus photographs.[3-5] In gastroenterology, AI is being tailored to address clinical questions and assist in medical decision making. It has utility in a broad spectrum of digestive diseases, with AI being developed for all subspecialties of gastroenterology. With the rapid expansion of diagnostic imaging and treatments available to gastroenterologists, the field is primed to reap the benefits of AI implementation. The following is an overview of AI as it applies to a critical aspect of the field—endoscopy.

DEVELOPMENT OF MACHINE LEARNING

Machine learning is a subset of AI. A computer can continue to learn based on information or experiences it has already learned. This provides the machine with an opportunity to improve its ability to perform a task over time without having been explicitly programmed for such a task. Hence, with machine learning, more complex functions and tasks can be learned, such as algorithms for self-driving cars and flying of drones.[6]

A common area of interest in AI development is facial recognition. DeepFace is a continually developing neural network used to identify faces in images. Trained by millions of social

media images, it reaches an accuracy of 97.35% on unclassified images and is bridging the gap between automated and human facial recognition.[7] DeepFace has been implemented in social media to identify faces and provide recommendations for users of the platform. This exact technology can be applied to endoscopic images where lesions are identified, differentiated, and classified similar to faces of various individuals through the use of convolutional neural networks (CNNs).

CONVOLUTIONAL NEURAL NETWORK

Advanced machine learning models such as CNNs are highly effective at image classification tasks. A CNN is designed to think similarly to the human brain, using large image datasets to learn patterns in correlating images. It achieves this with a construction similar to how biological neurons are interconnected. By connecting inputs through pattern recognition, a CNN can "learn" the process of classification much like a person. These models are trained with datasets containing images that have an element of interest versus datasets that do not. Typically, a dataset is randomly partitioned to reserve subsets for cross-validation.[8] This allows the developing model to experience new information and ensure that the model

does not "overfit" on previously seen data. Over time, CNNs learn how to detect the element of interest more accurately. A CNN makes inferences and predictions as if it believes that the element of interest is present within any given image, even if it has never seen that specific image before (Fig. 1).

The use of CNN is ideal for training an algorithm to detect and localize a lesion in endoscopy, such as detecting the presence of a polyp on colonoscopy screening. Using the images labeled by an observer, the algorithm has learned what an adenoma looks like on colonoscopy imaging. Subsequently, CNN is tested on images it had not previously seen to validate that the resulting model can identify previously unseen adenomas. The result is an algorithm that will place a bounding box around what it believes is likely to be an adenoma in a real-time colonoscopy video feed. Such a technology might significantly improve an endoscopist's adenoma detection rate (ADR), as we will discuss later.[5]

APPLICATIONS OF AI IN COLONOSCOPY

Colon polyp detection

Colorectal cancer remains the second most common cause of cancer-related deaths within the United States. The Amer-
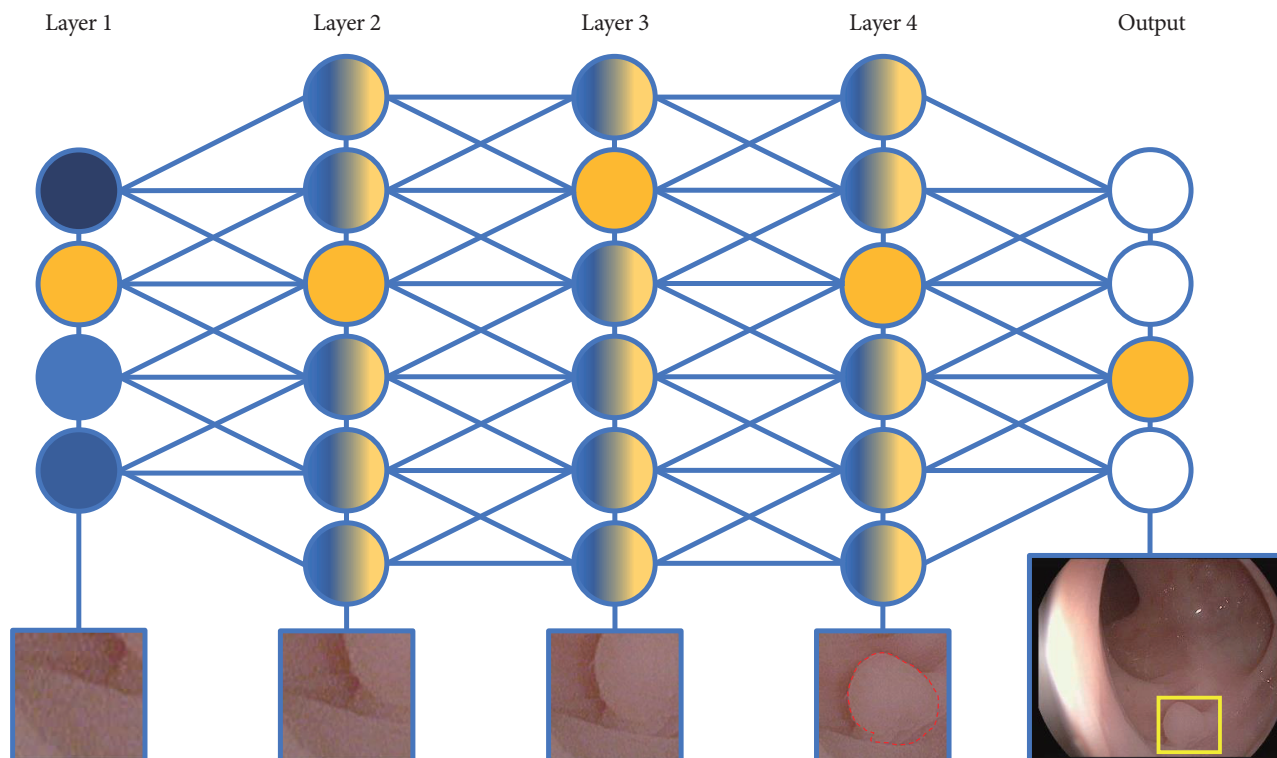
Fig. 1. Convolutional neural network (CNN) identification of a polyp. The CNN identifies the edges of a polyp via color and contour patterns. It further expands its recognition, going through multiple pathways to expand the identification area. The CNN identifies a pattern that it suspects as a polyp, tracing the borders of the lesion. Finally, the CNN reports the lesion to the endoscopist via a rectangular box around the polyp on the colonoscope output feed.

ican Cancer Society estimated there were 51,020 colorectal cancer deaths in the United States in 2019.[9] However, effective colorectal cancer screening has been shown to detect precancerous polyps early. With every increase in 1% of a gastroenterologist's ADR, the risk of an interval cancer decreases by 3% to 6%.[10,11] However, ADRs vary greatly between colonoscopists (from 7% to 53%).[10]

With the health risks of colon cancer, AI-assisted colon polyp detection has been an area of great interest. Multiple AI algorithms have been developed to run real-time during colonoscopy and will alert an endoscopist of the presence of polyps with either a visual cue or a sound. Karnes et al. developed an adenoma detection model that can effectively identify premalignant lesions.[12,13] Using images from 8,641 screening colonoscopies, a CNN was developed with 96.4% accuracy at a maximum rate of 170 images/s.[12,13] This CNN was shown to aid polyp detection, even for high-performing colonoscopists. Nine colonoscopy videos were reviewed by three expert colonoscopists (ADR ≥50%). One senior expert (ADR ≥50% and >20,000 colonoscopies) then reviewed all polyps found by the original colonoscopist, expert panel, and by the machine learning model and labeled each polyp as either low- or high-confidence for the presence of a polyp. In these videos, 28 polyps were initially identified by the original colonoscopists, which were removed. However, the expert review group identified a total of 36 polyps, while the CNN identified a total of 45 polyps without missing any unique polyps. On review, the CNN's additional 9 polyps were split between 3 high-confidence and 6 low-confidence polyp candidates. Machine learning vs. expert review of machine-overlaid videos had a sensitivity and specificity of 0.98 and 0.93, respectively (Chi-Square p<0.00001). By identifying 9 additional polyps compared with an expert panel, this CNN demonstrates the ability to assist even high-ADR colonoscopists during live colonoscopy.[5,14]

Additional research groups have studied polyp detection. Tajbakhsh et al. developed a computer-assisted diagnosis (CAD) system to detect polyps in real-time.[15] A hybrid context shape approach was used, in which non-polyp structures were first removed from a given image followed by localization of polyps by focusing on areas with curvy boundaries. This group reported an 88% sensitivity in real-time polyp detection.[15] Fernández-Esparrach et al. developed a CAD system to detect polyps by evaluating polyp boundaries and generating energy maps that corresponded to the presence of a polyp.[16] In that study, 24 videos containing 31 different polyps were included. The sensitivity and specificity for the detection of polyps were 70.4% and 72.4%, respectively.[16] Each study demonstrates a different approach to improving the effectiveness of colonoscopy screenings for identifying lesions of interest.

Colon polyp optical pathology

It has been shown that routine post-polypectomy pathology diagnosis and real-time endoscopic optical diagnosis are comparable amongst small colorectal polyps. Optical diagnosis may be more cost-effective and time-efficient compared with traditional post-polypectomy pathology diagnosis, as polyps do not need to be acknowledged or reviewed by a pathologist. For this reason, optical pathology has gained more acceptance. Multiple groups have worked toward developing CAD systems, in which computational analysis is used to predict polyp histology.[17]

Min et al. created a CAD system to predict adenomatous polyp vs. non-adenomatous polyp histology using linked color imaging.[18] This system achieved an accuracy of 78.4%, which is comparable to the accuracy of expert colonoscopists. The system had a sensitivity of 83.3%, specificity of 70.1%, positive predictive value (PPV) of 82.6%, and negative predictive value (NPV) of 71.2% in determining an adenomatous vs. non-adenomatous polyp.[18] With narrow-band imaging (NBI) magnification, Kominami et al. developed a CAD that achieved an accuracy of 93.2%, a sensitivity of 93.0%, a specificity of 93.3%, PPV of 93.0%, and NPV of 93.3% between colorectal polyps' histologic findings and real-time image recognition.[19] Aihara et al. developed a CAD model based on "real-time" color analysis of colorectal lesions using autofluorescence endoscopy.[20] The model had a sensitivity of 94.2%, specificity of 88.9%, PPV of 95.6%, and NPV of 85.2% in distinguishing neoplastic and non-neoplastic lesions during screening colonoscopies.[20] Renner et al. developed a CAD system to distinguish neoplastic vs. non-neoplastic polyps using unmagnified endoscopic images.[21] Optical pathology was compared with histopathological diagnoses. The optical pathology CAD system achieved an accuracy of 78%, sensitivity of 92.3%, and NPV of 88.2% compared with histopathological diagnoses.[21]

Zachariah et al. recently released the findings of a colon polyp AI with high optical pathology predictions.[22] Proposed in 2011 by the American Society of Gastrointestinal Endoscopy's "Preservation and Incorporation of Valuable Endoscopic Innovations", the goal is to achieve an accuracy of >90% and an NPV of >90% for optical pathology compared with traditional histopathology. These thresholds may allow the "resect and discard" strategy for accuracy and "diagnose and leave" for NPV of diminutive polyps of ≤5 mm. These small polyps typically dictate surveillance colonoscopy intervals in the current guidelines and are, therefore, of high interest for cancer prevention despite having 0.3% high-grade dysplasia. Using a large dataset of over 180,000 polyps, Zachariah et al. selected 5,278 high-quality images from typical white-light and NBI imaging systems of individual unique diminutive polyps with known pathology.[22] Of these images, 3,310 were adenomatous

polyps (tubular, tubulovillous, villous, and flat adenomas) versus 1,968 serrated polyps (hyperplastic and sessile serrated polyps) and were partitioned into 5 subsamples (80% for training and 20% for validation with cross-fold validation). With a focus on comparing with histopathology as the gold standard, the overall accuracy of optical pathology with CNN was 93.6%, with an NPV of 92.6%, PPV of 94.1%, sensitivity of 95.7%, and specificity of 89.9%. Both white-light and NBI were tested without significant difference in their performance. As overall accuracy was >90% and NPV was >90% (Fig. 2), this study demonstrated that optical pathology diagnosis of colorectal polyps using a CNN is feasible and may achieve more strict goals, which eliminates the need for a histopathological diagnosis by pathologists. Even more impressive, this CNN could run in real-time and did not require specialized imaging systems or light sources beyond what is commonly available in the gastrointestinal suite.[22]

Endocytoscopy (EC), using specialized endocytoscopes that have a forward-facing microscope capable of real-time magnification of over ×1,000, compared with the typical ×50 available with a traditional endoscope, EC has opened an entirely new imaging technique in gastroenterology, promising a diagnosis without resection. Several research groups have started to combine EC with CAD systems to generate advanced AI assistants, which have achieved in-procedure pathologic diagnosis. Mori et al. in 2015 developed the EC-CAD system, which used stained feature extraction to predict neoplastic polyps in 152 patients.[23] Polyps less than 10 mm were analyzed by EC in real-time for neoplastic changes. The CAD achieved

a sensitivity of 92.0% and specificity of 79.5%, with an accuracy of 89.2% for identifying neoplastic changes, comparable to those of expert endoscopists.[23] Removing the need for staining, Misawa et al. further refined the EC model using NBI to achieve more impressive results with the overall sensitivity of 84.5%, specificity of 97.6%, and accuracy of 90.0% using the existing training images.[24] When the resulting probability of diagnosis was >90%, the result was considered a "high-confidence" diagnosis. These diagnoses carried an overall sensitivity of 97.6%, specificity of 95.8%, and accuracy of 96.9%, surpassing the proposed cutoffs for a "diagnose-and-leave" system rather than universal resection with histopathologic diagnosis.[24]

AUTOMATED ENDOSCOPY REPORT WRITING

There has been great interest amongst our research group in the concept of report "auto-documentation" using AI. Computer vision AI algorithms can be used to "observe" the technical aspects of the procedure and then document the activity. The following is a summary of the work conducted to date for colonoscopy report writing.

Automated cecum detection

Karnes et al. developed a CNN to identify the cecum to develop automated cecal intubation and withdrawal time documentation.[25] Here, 6,487 colonic images throughout the entire

Fig. 2. Optical pathology algorithm of colon polyps showing an adenoma prediction (A) and serrated polyp prediction (B).

colon were annotated and labeled with anatomic landmarks, lesions, and preparation quality by two experienced colonoscopists. With both white-light and NBI images, a binary classification model was used to differentiate the cecum versus other colonic areas, with 80% of the images used for training and the remaining 20% for validation. In preliminary results for a smaller subset (1,000 images), the model achieved 98% accuracy and 0.996 area under the receiver operator characteristic curve. When the entire dataset was used, the accuracy decreased to 88%, which may have been due to more visually challenging images in the dataset. Further improvements in accuracy and validation of this model may help lead automation in the documentation of cecal intubation and calculation of total withdrawal time.[25]

Automated device recognition

Working toward automated recognition of endoscopic tools, Samarasena et al. developed a CNN to detect devices during endoscopy and colonoscopy.[26] Images of snares, forceps, argon plasma coagulation catheters, endoscopic assist devices, dissection caps, clips, dilation balloons, loops, injection needles, and rothnets were all studied by the CNN. A total of 180,000 images (56,000 of which had had at least one of these devices) were used to create the model. Further, 80% of the images were used to train the model and 20% were used for validation. The CNN had an accuracy of 0.97, area under the curve of 0.96, sensitivity was 0.95, and specificity was 0.97 for detecting these devices. Moreover, the system processed each frame within 4 ms. The algorithm has accurately detected these devices in a still image.[26]

Automated bowel preparation score management

Because of a strong correlation between the quality of colonoscopy preparation and ADR, bowel preparation score is another metric that should be automatically assessed with AI. The Boston bowel preparation scale (BBPS) is commonly used to assess bowel preparation quality but has a significant interoperator variability. To minimize interoperator variability, Karnes et al. developed a CNN to score bowel preparation quality.[27] A total of 3,843 colonic images were reviewed by two expert colonoscopists, who scored each bowel preparation image on a BBPS of 0–4. Then, 80% of these images were used to train the model and 20% were used to validate the model. The model had an accuracy of 97% when identifying a single frame as adequate (BBPS, 0–1) or inadequate (BBPS, 2–3) bowel preparation. The model processed each frame within 10 ms.[27]

Automated polyp size measurement

Accurate polyp size measurement is important for determining an appropriate timing of surveillance colonoscopy

screening. However, there is an inconsistency in accurately determining polyp size despite using rulers or biopsy forceps. Requa et al. developed a CNN to estimate polyp size during live colonoscopy.[28] A total of 8,257 images of polyps were included and labeled into three different size groups—diminutive (≤5 mm), small (6–9 mm), and large (≥10 mm)—by a single expert colonoscopist who has performed over 30,000 colonoscopies with an overall ADR of 50%. The resulting model had an accuracy of 0.97, 0.97, and 0.98 for polyps ≤5 mm, 6–9 mm, and ≥10 mm, respectively, and processed 100 fps, capable of being run during live colonoscopy. The ability to accurately categorize polyps into three different size groups in real-time colonoscopy may help endoscopists more accurately and consistently determine an appropriate timing of surveillance colonoscopy and additionally document polyp sizes during live colonoscopy.[28]

ARTIFICIAL INTELLIGENCE IN INFLAMMATORY BOWEL DISEASE

AI has already been applied to assisting in diagnosis of inflammatory bowel disease (IBD). Mossotto et al. describe using machine learning to develop new models for pediatric IBD classification.[29] They developed a model that incorporated both endoscopic and histologic data to separate pediatric IBD into Crohn's disease or ulcerative colitis with 82.7% accuracy, with the presence of ileal disease, the single most important factor for disease classification.[29] Maeda et al. used EC to examine histologic inflammation in ulcerative colitis.[30] With EC, they developed an AI model that identified persistent histologic inflammation with 74% sensitivity and 97% specificity. It accurately detected histologic inflammation in 90% of cases. The identification of inflammation had 100% interobserver reproducibility with the EC system.[30]

Historically, endoscopic scoring of IBD has been a challenge regarding inter and intra-observer agreement. Commonly used scores such as the endoscopic subscore or the Mayo Score have been challenged by a lack of clinical validation and a disagreement on repeated observations. Alternative scoring systems have been developed to correlate to clinical status or pathology findings yet rely on calculations and measurements that are more time-consuming.[31] "Central readers", clinically blinded endoscopic scorers who received specialized training, are an alternative approach to improve interobserver agreement in clinical trials.[32] AI scoring is an alternative to these models. The Mayo endoscopic subscore was used as the basis to develop a CNN scorer for ulcerative colitis. Abadir et al. have demonstrated that it is possible to create a CNN capable of strongly differentiating between mild and severe endoscop-

EARLY GASTRIC CANCER DETECTION

Gastric cancer is the third leading cause of cancer-related death worldwide. Gastric cancer can easily be missed on routine imaging and endoscopy, especially in countries with a low incidence of the disease and where training is limited. The 5-year survival rate of gastric cancer highly correlates with the stage of gastric cancer at the time of diagnosis. Therefore, it is paramount to improve our detection rates of early gastric cancer. Many groups have already started integrating AI into their routine practice to improve their overall detection rates of gastric cancer.[35] For endoscopists, gastric cancer has many visual features that are challenging to describe. To identify these subtle findings, Hirasawa and colleagues used 2,296 images (714 with confirmed gastric cancer) to develop a CNN with an overall sensitivity of 92.2% to detect gastric cancer via imaging alone.[35]

To localize blind spots during an esophagogastroduodenoscopy (EGD) that may have otherwise been missed, Wu et al. produced the WISENSE system, a real-time CNN to detect blind spots.[36] These blind spots of gastric mucosa, such as the lesser curvature of the antrum and the fundus, are areas that may occur depending on the endoscopist's abilities and may hide lesions. Trained on 34,513 images for gastric sites agreed on by at least four endoscopists, WISENSE detected blind spots with an accuracy of 90.02% by identifying these anatomic landmarks in EGD. In a single-center randomized control trial, 153 patients had their blind spots detected by WISENSE vs. 150 for the control group with unaided EGD. In the WISENSE group, the rate of blind spots was 5.86% versus 22.46% for the control group, a significant difference. This may be in part due to longer inspection times of an average 5.03 min in the WISENSE group vs. 4.24 min for the control group; the time could be longer in part because the AI has reminded the endoscopist to check the blind spots. WISENSE had the added benefit of providing more complete endoscopy imaging reporting than the control-group endoscopists.[36]

Magnified NBI (M-NBI) has been shown to have higher detection rates of early gastric cancer. However, many endoscopists are not trained to confidently use M-NBI. To address this, Kanesaka and colleagues developed a CAD to help diagnose early gastric cancer using only M-NBI images. The system that identifies gastric cancer using M-NBI achieved an accuracy of 96.3%, PPV of 98.3%, sensitivity of 96.7%, and specificity of 95%. Their CAD processed each image in 0.41 s.[37]

Helicobacter pylori is associated with an increased risk of gastric cancer but can be difficult to identify endoscopically. Therefore, an early detection has been an area of interest for the reduction in gastric cancer. Itoh and colleagues trained a CNN to detect the subtle endoscopic features caused by *H. pylori*. Their CNN achieved a sensitivity and specificity of 86.7% to detect *H. pylori* via a single endoscopic image.[38] Shichijo et al. developed a similar CNN to detect *H. pylori* within a single endoscopic image with a sensitivity, specificity, accuracy, and diagnostic processing time of 81.9%, 83.4%, 83.1%, and 198 s, respectively.[39] These findings are comparable to the detection rates of *H. pylori* by an endoscopist with the added benefit of identifying *H. pylori* much faster via computer assistance.[39]

Determining the depth of gastric cancer invasion is critical for prognosis. However, gastric cancer depth can be difficult to determine with endoscopy alone. Kubota et al. developed a computer-aided system to determine the depth of wall invasion from an endoscopic image.[40] The overall accuracy of the correct depth of invasion was 64.7%. The accuracy based on gastric wall invasion was 77.2%, 49.1%, 51.0%, and 55.3% for T1, T2, T3, and T4 staging, respectively. This computer-aided system demonstrates a novel approach to determine the depth of gastric cancer wall invasion via endoscopy.[40]

Regarding diagnosis, it is vital to have an accurate and consistent histological interpretation of a prepared gastric cancer slide. To minimize the intra- and interobserver variability in interpreting a gastric biopsy, Sharma and colleagues developed a CNN. Their CNN was trained to interpret histological images of gastric cancer, achieving an accuracy of 69.9% for gastric cancer identification and 81.4% for necrosis detection. Further development of a similar CNN can lead to a more standardized, consistent, and accurate diagnosis of gastric cancer.[41]

ESOPHAGEAL NEOPLASIA DETECTION

Barrett's esophagus is a topic of intensive investigation for AI development. Traditionally difficult to detect with imaging, serial biopsies have been the gold standard for the identification of this esophageal pathology. However, with the subtle visual changes present in Barrett's esophagus with dysplasia, there is a considerable risk of missing a diagnosis of dysplasia on endoscopy. Among expert and community endoscopists, the rates of dysplasia detection (both with lesions that could be visually identified or not) remain highly variable.[42] This

provides an opportunity for AI systems to improve patient care and early diagnosis of esophageal precancerous lesions. Van der Sommen et al. worked toward developing a CAD from high-definition images to detect dysplastic lesions in Barrett's esophagus.[43] This system was constructed with 100 annotated endoscopic images and compared with expert scorers. During the final analysis, the expert scorers demonstrated consistently superior sensitivity and specificity (greater than 95% sensitivity and 65%–91% specificity) compared with the CAD systems tested, and the system demonstrated per-image sensitivity and specificity of 83% for dysplastic lesions. While not compared with nonexperts, this early feasibility study demonstrated that, with sufficient training, a CAD system may eventually reach the American Society of Gastrointestinal Endoscopy's target of per-patient sensitivity of 90% or higher for an optical imaging technology.[43]

De Groof et al. continued work on CAD development for Barrett's esophagus detection.[44] Prospective collection of white-light imaging from 40 Barrett's and 20 non-dysplastic patients were delineated by six experts for suspected areas of Barrett's. Areas with >50% overlap by the experts were the "sweet spot" and with one expert identifying dysplasia the "soft spot" used to train the CAD. Per-image, an accuracy of 91.7%, sensitivity of 95%, and specificity 85% was achieved by the CAD. Localization was successful in 100% of the soft spots and 97.4% of the sweet spots, indicating substantial agreement with the experts. Finally, "red flags" indicating CAD-identified dysplasia were placed within 89.5% of soft spots and 76.3% of sweet spots. With a mean time of 1.051 s/image, the algorithm rapidly identified Barrett's esophagus in white-light imaging.[44]

Our research group described the development of a high-functioning CNN to identify early neoplasia within Barrett's esophagus. Trained on 916 images of early esophageal

Fig. 3. Optical pathology algorithm predicting an area of dysplasia within a segment of Barrett's esophagus.

neoplasia or T1 adenocarcinoma confirmed by biopsy from 65 patients, this image set was combined with an equal number of Barrett's esophagus without high-grade dysplasia to create a training set of images, with 458 images separated for validation. The CNN correctly identified neoplasia in Barrett's esophagus with a sensitivity of 96.4%, specificity of 94.2%, and accuracy of 95.4% (Fig. 3). Localization of the dysplasia was also highly accurate. Furthermore, the speed at which predictions were made was well above 70 fps. Hence, it is possible to develop a CNN system for dysplasia detection during real-time endoscopy.[45]

For esophageal cancer detection, Horie et al. used 8,420 images of esophageal cancer from 384 patients to develop a CNN based on the single-shot multibox detector framework.[46] Using images from white-light and NBI, the constructed CNN identified all lesions over 10 mm with an overall sensitivity for esophageal cancer detection of 98%. The system analyzed 30 images/s.[46]

Cai et al. also developed a system to identify early esophageal squamous cell carcinoma.[47] Using 2,428 standard EGD white-light images from 746 patients, 1,332 showed abnormal endoscopic tissue. Endoscopic submucosal dissection was performed on all cases with abnormal visual findings to verify pathology, while 3 experienced endoscopists annotated the images for pathology based on NBI images and pathology results (2 classifying with 1 verifying). A validation set of 187 images, reviewed by blinded endoscopists separated by seniority, was used to compare the results with the 8-layer CNN. The CNN had a sensitivity, specificity, accuracy, PPV, and NPV of 97.8%, 85.4%, 91.4%, 86.4%, and 97.6% with an AUC of 0.96 for early esophageal squamous cell carcinoma. When compared with the blinded endoscopists, the CNN had a greater sensitivity, accuracy, and NPV than any of the endoscopists. However, all groups demonstrated improvement in metrics when referring to the CNN results, demonstrating that the CNN may assist the endoscopists in diagnosis.[47]

CAPSULE ENDOSCOPY

AI use in capsule endoscopy is a growing research field. Capsule endoscopies are often multiple hours long with a significant time burden for the reader. Moreover, the same study may occasionally miss a pathology, as confirmed by a review of a second reader.[48] There have been programs developed to improve the speed at which a reader can assess a complete capsule study, but they have been shown to occasionally miss a significant pathology.[49] With large amounts of video data, capsule studies are an ideal target of AI development to aid gastroenterologists in identifying distinct landmarks and areas
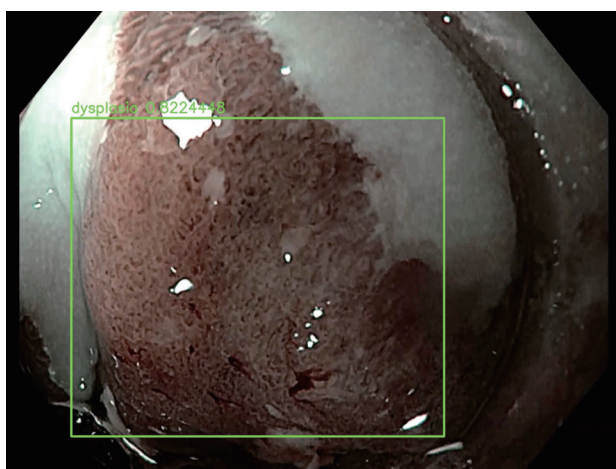
of interest. As newer generations of high-definition capsules emerge, richer data are available for training a CNN to detect masses or source of occult bleeding. Leenhardt et al. developed a CNN to detect gastrointestinal angioectasias in the small bowel.[50] This model is proved to have 100% sensitivity for detecting angioectasias, with a specificity of 96%. This algorithm assesses a full-length study in 39 min, with 46.8 s/still frame.[50]

Recently, Lui et al. presented data acquired from 439 capsule videos used to create a CNN that identified multiple types of lesions and their lumen locations.[51] With an accuracy of 97%, the model identified arteriovenous malformations, erythema, varices, bulges, masses, ulcerations, erosions, blood, red villi, diverticula, polyps, and xanthomas on capsule endoscopy images. It achieved 98% accuracy for location identification, and both models processed an entire capsule study in 10–15 min at 500 fps.[51]

Ding et al. published results on a small bowel capsule CNN, which was developed using data from 77 medical centers supplying 113 million images from 6,970 patients.[52] Normal small bowel images were distinguished from lesions of interest in the initial 1,970 patients to develop the CNN with the other 5,000 patients used for validation. CNN predictions were compared with conventional capsule analysis by trained gastroenterologists to confirm or reject CNN categorization. The resulting 4,206 lesions of interest were identified by the CNN with 99.88% sensitivity in a per-patient analysis, compared with 74.57% sensitivity by gastroenterologists. Mean reading time by the CNN was 5.9 min compared with 96.6 min by the traditional method, demonstrating the impressive efficiency of CNNs in capsule endoscopy.[52]

Prospective trials are necessary for verifying the promising results of CNNs in capsule endoscopy.[53] However, when such CNNs are ready for clinical prime-time, they are expected to revolutionize the practice of capsule endoscopy through drastically improved detection rates and efficiency.

LIMITATIONS OF AI USE

AI has not yet been used in large-scale clinical applications or patient care outside of limited clinical research studies. Likewise, as of now, there are no AI algorithms approved for use in clinical gastroenterology by governmental organizations for standard patient care. The difficulty for AI integration into common practice is multifactorial, with major obstacles prior to large-scale implementation.

First, the development of AI requires a human diagnosis or image label to train the AI. This builds an inherent bias into the system from the start, as the AI is "learning" from en-

doscopists who may have personal thresholds for classifying findings. The worst example of this is an AI developed with the input of a single, nonblinded endoscopist, while studies have attempted to avoid this bias by blinding or having findings go through multiple endoscopists to classify them, with additional endoscopists acting as "tie breakers" when there is discordance amongst classifications.

As most CAD or CNN systems have been developed in single centers, there may be overarching institutional diagnostic or selection bias, which may complicate more widespread implementation. As each institution services a unique patient population, having AI development in single centers raises the concern of spectrum bias, when the AI is taken out of the original clinical context. Its behavior may be unpredictable and it may not maintain the same diagnostic accuracy as before. To adopt any of the current-generation AIs into common practice, it will have to be shown that they consistently perform in a variety of clinical locations and scenarios (such as different endoscopists, endoscopy equipment, clinical software, and other ancillary staff that may play a role in endoscopy). This will require substantial external clinical validation beyond the original clinical centers of development for each AI prior to more widespread adoption, which current AI developers have not attempted yet.

On the topic of validation, it is difficult to perform clinical trials, especially blinded randomized-control trials, with AI. These types of trials are the most robust and the "gold standard" in medical science. While prospective trials have been performed, this difficulty with initiating more robust studies has resulted in most AI systems for endoscopy being developed retrospectively, nonrandomized, with limited blinding by the clinicians.[54] This limits practical adoption of AI by clinicians and professional organizations due to the overall weaker body of data supporting each AI endoscopy system.

When AI systems are developed, current AI interfaces are research-focused and not adapted to everyday use. Some, such as those using EC, require additional equipment. However, there are research groups attempting to create easy-to-use technologies that may be implemented with more common endoscopes and software systems.[22] However, the implementation and use of AI require additional training of endoscopists and the support staff. This includes an understanding of the false positives that CNNs often identify on endoscopy, particularly when there are visual hurdles, such as poor bowel preparation or excessive bubbles in imaging. While several AIs are being developed for automatic report writing, they will still need to communicate effectively with electronic record systems to ensure efficient documentation.

As current-generation AIs are numerous but proprietary, the cost may be a restrictive factor for widespread implemen-

tation. The final hurdle will be the acceptance and approval of AI by governmental and professional organizations. To authorize specific systems for regular patient care, the previously described limitations should be mitigated prior to approval by health ministries and other regulatory authorities. The professional organizations and the general public will have their concerns regarding the implementation of such a novel technology into the endoscopic diagnosis, which may replace more traditional diagnostic techniques. For each country, any legal question regarding the use of AI in endoscopy (such as AI misdiagnosis) will have to be addressed prior to widespread implementation. Then, usage guidelines must be established by gastroenterological organizations to ensure that specific AIs are being used in appropriate situations. While this is a long process toward clinical use of AI-assisted endoscopy, it has already begun in many countries where AI development is ongoing.

CONCLUSIONS

AI has made a dramatic entrance into the field of gastroenterology. AI models are being developed constantly by academic and industry research groups to provide objective diagnostic support to endoscopists. As in medicine as a whole, AI is creating a future in gastroenterology where the provider relies upon trained AI tools for diagnosis and feature identification, reducing the occurrence of missed diagnoses and lesions, which may impact patient care. With automated report-writing, CNNs may provide dramatically improved efficiency and accuracy to procedural documentation.

A unique feature of CNNs and other models is their ability to improve themselves continually as more data are provided. Hence, a future is possible in which researchers and manufacturers would use the provider's data to improve models continuously through regular upgrades. However, this feature will not remove the intrinsic limitations of CNNs and other AI models; therefore, practitioners will have to remain aware of these limitations during the AI usage.

Supervised clinical implementation of AI models will be necessary to assess their real-world use in a large-scale setting. Prospective randomized control trials would be beneficial in verifying the largely retrospective studies used for model creation. However, as AI further permeates the medical specialties, it is expected that AI will gain wider acceptance as a clinical tool in everyday practice. While it remains to be seen how exactly and to what extent AI will be implemented in gastroenterology subspecialties, it is predicted to be a monumental shift in practice that will substantially elevate patient care.

ORCID

Alexander P. Abadir: https://orcid.org/0000-0002-5187-5128
Mohammed Fahad Ali: https://orcid.org/0000-0002-1181-4902
William Karnes: https://orcid.org/0000-0002-6225-9080

REFERENCES

1. Alagappan M, Brown JRG, Mori Y, Berzin TM. Artificial intelligence in gastrointestinal endoscopy: the future is almost here. World J Gastrointest Endosc 2018;10:239-249.
2. Silver D, Huang A, Maddison CJ, et al. Mastering the game of Go with deep neural networks and tree search. Nature 2016;529:484-489.
3. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542:115-118.
4. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016;316:2402-2410.
5. Urban G, Tripathi P, Alkayali T, et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. Gastroenterology 2018;155:1069-1078.e8.
6. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436-444.
7. Taigman Y, Yang M, Ranzato M, Wolf L. Deepface: closing the gap to human-level performance in face verification. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition; 2014 Jun 23-28; Columbus (OH), USA. Piscataway (NJ): IEEE; 2014. p. 1701-1708.
8. Ruffle JK, Farmer AD, Aziz Q. Artificial intelligence-assisted gastroenterology- promises and pitfalls. Am J Gastroenterol 2019;114:422-428.
9. National Cancer Institute. Cancer stat facts: common cancer sites [Internet]. Bethesda (MD): NCI; c2019 [cited 2020 Feb 3]. Available from: https://seer.cancer.gov/statfacts/html/common.html.
10. Corley DA, Jensen CD, Marks AR, et al. Adenoma detection rate and risk of colorectal cancer and death. N Engl J Med 2014;370:1298-1306.
11. Kaminski MF, Wieszczy P, Rupinski M, et al. Increased rate of adenoma detection associates with reduced risk of colorectal cancer and death. Gastroenterology 2017;153:98-105.
12. Karnes WE, Alkayali T, Mittal M, et al. Automated polyp detection using deep learning: leveling the field. Gastrointest Endosc 2017;85(5 Suppl):AB376-AB377.
13. Karnes WE, Ninh A, Urban G, Baldi P. Adenoma detection through deep learning: 2017 presidentialposter award. Am J Gastroenterol 2017;112:S136.
14. Tripathi PV, Urban G, Alkayali T, et al. Computer-assisted polyp detection identifies all polyps found by expert colonoscopists – and then some. Gastroenterology 2018;154(6 Suppl 1):S-36.
15. Tajbakhsh N, Gurudu SR, Liang J. Automated polyp detection in colonoscopy videos using shape and context information. IEEE Trans Med Imaging 2016;35:630-644.
16. Fernández-Esparrach G, Bernal J, López-Cerón M, et al. Exploring the clinical potential of an automatic colonic polyp detection method based on the creation of energy maps. Endoscopy 2016;48:837-842.
17. Puig I, Kaltenbach T. Optical diagnosis for colorectal polyps: a useful technique now or in the future? Gut Liver 2018;12:385-392.
18. Min M, Su S, He W, Bi Y, Ma Z, Liu Y. Computer-aided diagnosis of colorectal polyps using linked color imaging colonoscopy to predict histology. Sci Rep 2019;9:2881.
19. Kominami Y, Yoshida S, Tanaka S, et al. Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. Gastrointest Endosc 2016;83:643-649.

20. Aihara H, Saito S, Inomata H, et al. Computer-aided diagnosis of neoplastic colorectal lesions using 'real-time' numerical color analysis during autofluorescence endoscopy. Eur J Gastroenterol Hepatol 2013;25:488-494.

21. Renner J, Phlipsen H, Haller B, et al. Optical classification of neoplastic colorectal polyps - a computer-assisted approach (the COACH study). Scand J Gastroenterol 2018;53:1100-1106.

22. Zachariah R, Ninh A, Dao T, Requa J, Karnes W. Can artificial intelligence (AI) achieve real-time 'resect and discard' thresholds independently of device or operator? Am J Gastroenterol 2018;113:S129.

23. Mori Y, Kudo SE, Wakamura K, et al. Novel computer-aided diagnostic system for colorectal lesions by using endocytoscopy (with videos). Gastrointest Endosc 2015;81:621-629.

24. Misawa M, Kudo S-E, Mori Y, et al. Characterization of colorectal lesions using a computer-aided diagnostic system for narrow-band imaging endocytoscopy. Gastroenterology 2016;150:1531–1532.e3.

25. Karnes WE, Ninh A, Dao T, Requa J, Samarasena JB. Real-time identification of anatomic landmarks during colonoscopy using deep learning. Gastrointest Endosc 2018;87(6 Suppl):AB252.

26. Samarasena J, Yu AR, Torralba EJ, et al. Artificial intelligence can accurately detect tools used during colonoscopy: another step forward toward autonomous report writing: presidential poster award. Am J Gastroenterol 2018;113:S619-S620.

27. Karnes WE, Ninh A, Dao T, Requa J, Samarasena JB. Unambiguous real-time scoring of bowel preparation using artificial intelligence. Gastrointest Endosc 2018;87(6 Suppl):AB258.

28. Requa J, Dao T, Ninh A, Karnes W. Can a convolutional neural network solve the polyp size dilemma? Category award (colorectal cancer prevention) presidential poster award. Am J Gastroenterol 2018;113:S158.

29. Mossotto E, Ashton JJ, Coelho T, Beattie RM, MacArthur BD, Ennis S. Classification of paediatric inflammatory bowel disease using machine learning. Sci Rep 2017;7:2427.

30. Maeda Y, Kudo SE, Mori Y, et al. Fully automated diagnostic system with artificial intelligence using endocytoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with video). Gastrointest Endosc 2019;89:408-415.

31. Paine ER. Colonoscopic evaluation in ulcerative colitis. Gastroenterol Rep (Oxf) 2014;2:161-168.

32. Panés J, Feagan BG, Hussain F, Levesque BG, Travis SP. Central endoscopy reading in inflammatory bowel diseases. J Crohns Colitis 2016;10(Suppl 2):S542-S547.

33. Abadir AP, Requa J, Ninh A, Karnes W, Mattar M. Unambiguous real-time endoscopic scoring of ulcerative colitis using a convolutional neural network. Am J Gastroenterol 2018;113:S349.

34. Bossuyt P, Vermeire S, Bisschops R. Scoring endoscopic disease activity in IBD: artificial intelligence sees more and better than we do. Gut 2020;69:788-789.

35. Hirasawa T, Aoyama K, Tanimoto T, et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. Gastric Cancer 2018;21:653-660.

36. Wu L, Zhang J, Zhou W, et al. Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy. Gut 2019;68:2161-2169.

37. Kanesaka T, Lee TC, Uedo N, et al. Computer-aided diagnosis for identifying and delineating early gastric cancers in magnifying narrow-band imaging. Gastrointest Endosc 2018;87:1339-1344.

38. Itoh T, Kawahira H, Nakashima H, Yata N. Deep learning analyzes Helicobacter pylori infection by upper gastrointestinal endoscopy images. Endosc Int Open 2018;6:E139-E144.

39. Shichijo S, Nomura S, Aoyama K, et al. Application of convolutional neural networks in the diagnosis of Helicobacter pylori infection based on endoscopic images. EBioMedicine 2017;25:106-111.

40. Kubota K, Kuroda J, Yoshida M, Ohta K, Kitajima M. Medical image analysis: computer-aided diagnosis of gastric cancer invasion on endoscopic images. Surg Endosc 2012;26:1485-1489.

41. Sharma H, Zerbe N, Klempert I, Hellwich O, Hufnagl P. Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology. Comput Med Imaging Graph 2017;61:2-13.

42. Schölvinck DW, van der Meulen K, Bergman J, Weusten B. Detection of lesions in dysplastic Barrett's esophagus by community and expert endoscopists. Endoscopy 2017;49:113-120.

43. van der Sommen F, Zinger S, Curvers WL, et al. Computer-aided detection of early neoplastic lesions in Barrett's esophagus. Endoscopy 2016;48:617-624.

44. de Groof J, van der Sommen F, van der Putten J, et al. The Argos project: the development of a computer-aided detection system to improve detection of Barrett's neoplasia on white light endoscopy. United European Gastroenterol J 2019;7:538-547.

45. Hashimoto R, Lugo M, Mai D, et al. Artificial intelligence dysplasia detection (Aidd) algorithm for Barrett's esophagus. Gastrointest Endosc 2019;89(6 Suppl):AB99-AB100.

46. Horie Y, Yoshio T, Aoyama K, et al. Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. Gastrointest Endosc 2019;89:25-32.

47. Cai SL, Li B, Tan WM, et al. Using a deep learning system in endoscopy for screening of early esophageal squamous cell carcinoma (with video). Gastrointest Endosc 2019;90:745-753.e2.

48. Byrne MF, Donnellan F. Artificial intelligence and capsule endoscopy: is the truly "smart" capsule nearly here? Gastrointest Endosc 2019;89:195-197.

49. Shiotani A, Honda K, Kawakami M, et al. Analysis of small-bowel capsule endoscopy reading by using quickview mode: training assistants for reading may produce a high diagnostic yield and save time for physicians. J Clin Gastroenterol 2012;46:e92-e95.

50. Leenhardt R, Vasseur P, Li C, et al. A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. Gastrointest Endosc 2019;89:189-194.

51. Lui F, Rusconi-Rodrigues Y, Ninh A, Requa J, Karnes W. Highly sensitive and specific identification of anatomical landmarks and mucosal abnormalities in video capsule endoscopy with convolutional neural networks. Am J Gastroenterol 2018;113:S670-S671.

52. Ding Z, Shi H, Zhang H, et al. Gastroenterologist-level identification of small-bowel diseases and normal variants by capsule endoscopy using a deep-learning model. Gastroenterology 2019;157:1044-1054.e5.

53. Hwang Y, Park J, Lim YJ, Chun HJ. Application of artificial intelligence in capsule endoscopy: where are we now? Clin Endosc 2018;51:547-551.

54. Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. Gut 2019;68:1813-1819.